

An Improved K - Mode Algorithm for Facial Expression Image Clustering

Xibin Jia, Jianming Yuan* and Yujie Xiao

Faculty of Information Technology, Beijing University of Technology, Beijing, China

*Corresponding author

Abstract—When recognizing facial expression sequences by discrete Hidden Markov Model, it is necessary to cluster the image frames into several observation states. Considering the complexity of face images in emotion expressing, we partition the whole face into several sub-regions to do clustering separately. Then the categorical values from clustering results of each sub-region are combined to do the further clustering. In the paper, we propose an improve K-mode clustering algorithm called K-frequency to do the category data clustering. Instead of using a simple 0-1 method to determine the similarity between different samples, the K-frequency statistics the frequency of each categorical values of an image in a given cluster and sums them as the similarity between a sample with this cluster. Experiment results on CK+ database show that the bi-level clustering method with K-frequency algorithm outperforms the Single-layer clustering using K-means and bi-level clustering with K-mode. The improved k-modes algorithm K-frequency is more robust than the original k-mode algorithm in dealing with isolated points.

Keywords- cluster; categorical values; expression recognition

I. INTRODUCTION

Facial expression recognition is a technique that determines people's mood or thought activity of a person by analyzing the face expression and change of human face using computer. Facial expression recognition is conducive to a more intelligent human-computer interaction environment. It is widely used in distance education, safe driving, public surveillance, polygraph technology, computer games, clinical medicine, psychopathology analysis and psychology. In recent years, due to the development of video surveillance, especially in the field of security, video information is growing rapidly. How to quickly extract useful information from the cumbersome video data has become a new research hotspot.

In order to effectively extract the dynamic information in the expression sequence, the discrete Hidden Markov Model is a good choice[1][2]. But there is a problem that the number of observation states can't be too much, or it will exceed the memory capacity. We adopt the clustering algorithm to solve this problem.

Clustering analysis is a process of dividing the data collection into several subsets based on the similarity of each element. Clustering algorithm can be divided into the following categories: partition-based method, hierarchical method, density-based method, grid-based method, model-based method, fuzzy clustering method, graph theory method, fractal ,

Clustering method of complex network, bionics method, kernel clustering method, and many synthesis of these methods[3].

K-means is the most typical partition-based approach, first proposed by Macqueen in 1967[4].As the most classical clustering method, K-means is able to run efficiently and easy to accomplish. It works well for many issues. However, K-means is sensitive to the initial value of the cluster center, a small amount of noise or isolated points can have a significant impact on the average. In addition, K-means fail to deal with categorical values.

Since K-means can only deal with numeric values, many people have improved it. In Especially the K-modes algorithm proposed by Huang in 1998[5], which extend the K-means algorithm on categorical values. K-modes use a simple 0-1 method to determine the dissimilarity between different samples, but this method ignores the similarity between the character values in the class and weaken the internal similarity of the class.

In the paper, we propose an improve K-mode clustering algorithm called K-frequency to do the category data clustering. When evaluating which cluster a new sample belongs to, firstly, the algorithm statistics all categorical values' frequency of every cluster's dimension, in which the numerator is the times a categorical value of a cluster's dimension appears and the denominator is the number of samples of the cluster. Then the algorithm the frequencies from the first step whose categorical values are equal to the new sample in every dimension. Plus these frequencies, and a larger value indicates a higher similarity. That is to say ,this algorithm evaluate the similarity of a new sample to a cluster depends on the frequency sum of the corresponding categorical values.

When dealing with the issue of feature extraction of facial expression, in order to eliminate irrelevant information, we need to split off the unrelated areas, and leave only the sub-regions with large impact on the expression. In this situation, two clustering strategies are proposed in this paper. One is to combine the features of these sub-regions, and then cluster with K-means method directly .The another is to cluster each sub-region with K-means separately, then combine categorical values from clustering results of each sub-region, and finally cluster combined categorical values with K-modes or K-frequency method. We analyze the characteristics of cohesion, externality, isolation and so on though experiments in the CK+ database[6]: 1) When dealing with multiple sub-regions, although the Single-layer clustering using K-means works

better than the bi-level clustering method in cohesion[7], the separation[7] of the bi-level clustering method is better. 2) For the same K-modes and K-frequency algorithm, the cohesion of the latter is better, more robust to isolated points.

II. CLUSTERING METHOD

A. The Classical Clustering Method of Categorical Values K-modes

What value is called Categorical values? The value of Categorical values is a tag with a category attribute. It can't be compared in the digital sense. For example, the size of alphabet a and b can't be compared. Let nonempty finite set $X = \{x_1, x_2, \dots, x_N\}$ be the sample set which is composed of N samples. The sample $x_n = \{x_{n1}, x_{n2}, \dots, x_{nM}\}$ is composed of M attributes, where $x_n \in X$. The range of each attribute $A = \{a_1, a_2, \dots, a_T\}$ is composed of T categorical values and $x_{nm} \in A$.

The typical method for categorical values clustering is K-modes. Based on the partition-based principle, K initial clusters are set first and the cluster centers are define as $z_k = \{z_{k1}, z_{k2}, \dots, z_{kM}\}, k = 1, 2, \dots, K$. q_{kmt} represents the time that categorical value t occurs in the m-th attribute of the cluster k. The categorical value with the highest time is used as the representative value of attribute z_{km} :

$$z_{km} = a_{\underset{t}{\arg \max}(q_{kmt})} \quad (1)$$

The distance between a sample and the cluster center is defined as:

$$d(x_n, z_k) = \sum_{m=1}^M \delta(x_{nm}, z_{km}) \quad (2)$$

where δ is 0-1 variable, as shown in the following formula:

$$\delta(x_{nm}, z_{km}) = \begin{cases} 0, & x_{nm} = z_{km} \\ 1, & x_{nm} \neq z_{km} \end{cases} \quad (3)$$

As can be seen from (2), the smaller the distance value $d(x_n, z_k)$, the higher the similarity between the sample x_n and the cluster center z_k . When $d(x_n, z_k) = 0$, it indicates that the sample x_n is exactly the same as the class center z_k . The cluster process divides x_n into the cluster with the highest similarity and updates the clustering center.

K-modes algorithm steps are shown as follows:

Step 1: Initialize. K samples are randomly selected as initial samples for each cluster.

Step 2: Update the cluster center. Each attribute of a cluster center is taken as the final categorical value with the highest time of occurrences.

Step 3: Divide the subsequent samples. The new sample is added to the cluster with the highest similarity. Repeat step 2 and 3 until no new samples appear.

B. K-frequency Clustering Method

Although the K-modes algorithm solves the problem of categorical values clustering, the drawback is obvious. The K-modes algorithm only selects the categorical value with the largest times as the attribute value of the cluster center, ignoring the effect of other categorical values in the cluster on the distance. In particular, when the time of occurrences of two categorical values are all the maximum, or when the maximum time is very close to the next largest time, the cluster center directly ignores the other maximum time or the next largest time. That results in a greater deviation from the distance judgment

In order to solve this problem, this paper improved the cluster center generation algorithm. The clustering centers of each cluster are no longer composed of representative categorical values. Instead, the frequency of each categorical value in each attribute is counted to form a two-dimensional matrix $p_k, k = 1, 2, \dots, K$:

$$p_k = \begin{bmatrix} p_{k11} & p_{k12} & \dots & p_{k1M} \\ p_{k21} & p_{k22} & \dots & p_{k2M} \\ \dots & \dots & \dots & \dots \\ p_{kT1} & p_{kT2} & \dots & p_{kTM} \end{bmatrix} \quad (4)$$

where $\sum_{t=1}^T p_{kmt} = 1, t = 1, 2, \dots, T, m = 1, 2, \dots, M$.

The distance between a sample and the cluster center is defined as:

$$d(x_n, p_k) = \sum_{m=1}^M p_{kx_{nm}m} \quad (5)$$

The frequency $p_{kx_{nm}m}$ is equivalent to the time of occurrences of the categorical value x_{nm} under the m-th attribute dividing the number of occurrences of all categorical values under the current attribute.

The larger the value of $d(x_n, p_k)$, the higher the similarity between sample x_n and cluster center p_k . The

cluster process divides x_n into the cluster with the highest similarity and updates the clustering center.

The algorithm steps are as follows:

Step 1: Initialize. K samples are randomly selected as initial samples for each cluster.

Step 2: Update the cluster center. The frequency of each categorical value under each attribute is counted as the cluster center.

Step 3: Divide the subsequent samples. The new sample is added to the cluster with the highest similarity. Repeat steps 2 and 3 until no new samples appear.

III. CLUSTERING STRATEGY IN FACIAL EXPRESSION IMAGE CLUSTERING

People's expression is a dynamic process. In order to use the dynamic information of the expression, the discrete Hidden Markov Model (DHMM) is used as the classifier. Hidden Markov model is a double stochastic process, containing two states: hidden state and observation state. The hidden state that can't be directly observed, can be understood as an intermediate state of expression action. The observe state that can be observed directly, can be understood as an image of the expression sequence. Mutual transfer relations exist between hidden states, but there is no mutual transfer relationship between observation states. There is a one-way transfer relationship between the hidden state and the observed state. Observe the state requirement is a limited set of state. Due to the limitation of hardware cost and time cost, the original image needs to be reduced.

Two clustering strategies were designed based on the different trait of numerical and categorical values clustering. Before the feature extraction, in order to eliminate the irrelevant information, the unrelated areas are usually split off, and the areas that have a large influence on the expression are reserved. Due to facial gestures are focused on the areas around the eyes, nose and mouth, so only need to extract those areas to realize the expression recognition. The clustering operation can be executed both in sub-regions and the combined region, but if clustering in sub-regions, there will be need categorical values clustering operation after it, as shown in Figure I.

In Figure I(A), six sub-regions are extracted from each image of the expression sequence. Then they are combined and clustered using the K-means.

In Figure I(B), in contrast to Figure I(A), the same sub-regions is clustered using the K-means before combination. The combined categorical value is a 6-dimensional vector, which is then clustered using K-modes or K-frequency.

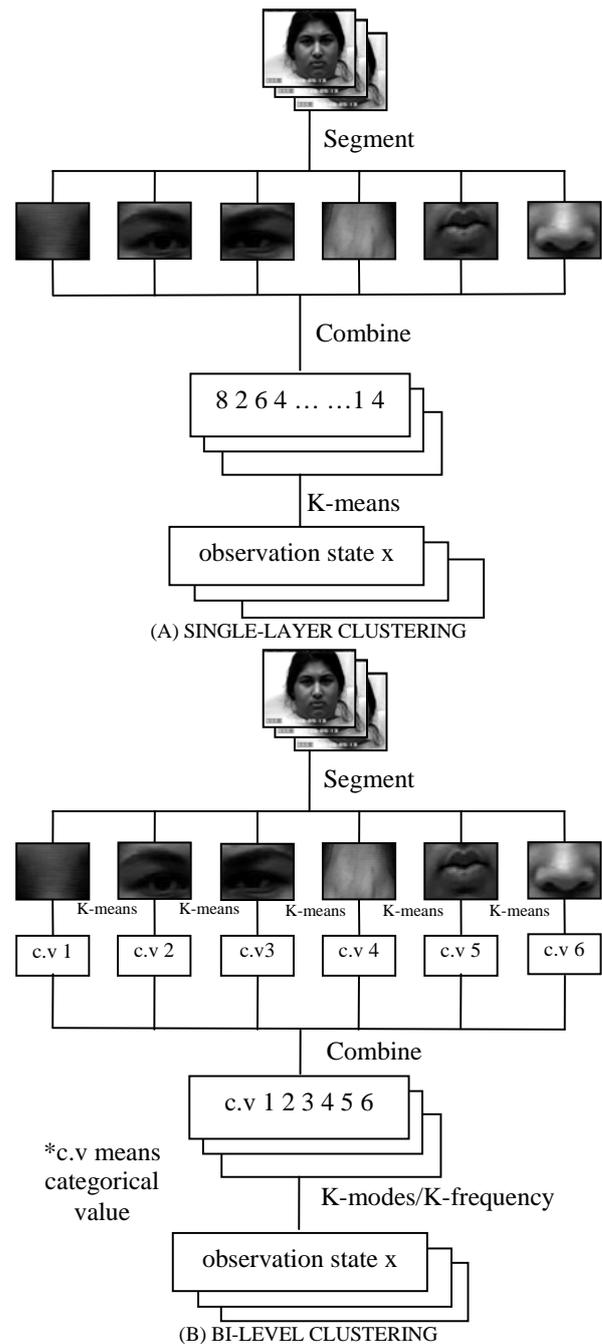


FIGURE I. CLUSTERING STRATEGY IN EXPRESSION SEQUENCE RECOGNITION

IV. EXPERIMENT AND RESULT ANALYSIS

In this study, we selected 327 expression sequences in CK + database, including 7 basic expressions (neutral, anger, contempt, disgust, fear, happiness, sadness, surprise). Data preprocessing is required before clustering, including sequence length normalization, face location and segmentation, and feature extraction.

First, the expression sequence is normalized to 20 frames. Thus the total number of images is 6540. Since 20 images share an emotion label, and the images under the same label differ in the degree of expression, the sequence labels do not match the image labels. So in cluster evaluation, only the internal validity index can be used. External validity index can't be used.

Using Face Tracker algorithm, forehead, left eye, right eye, between eyes, nose and mouth were extracted from the human face. Each feature was extracted using Gabor feature, and PCA dimensionality reduction was needed because of high feature dimension.

When the evaluation index is used for comparison, the clustering results of all the strategies are input in the form of "all features + label", so that the comparability between Euclidean distance and label-based clustering is realized. In the following, three kinds of clustering methods are evaluated by using seven kinds of indexes respectively.

TABLE I. BASIC INDICATORS

Cluster methods	within-class	among-class
K-means	1.21	2.88
K-modes	4.73	3.24
K-frequency	1.76	3.41

Basic indexes include within-class and among-class distance. The within-class reflects the compactness of cluster data. The smaller the within-class distance is, the higher the compactness of the cluster is, and the better clustering result is. The within-class distance of K-means in Table I is smaller than K-modes and K-frequency. This is because K-means does not involve the physical meaning of different feature segments, just the dividing data in the numerical sense, while within-class distance is calculated by numerical calculation. The disadvantage is that, due merged 6 features of each block before clustering, the proportion of those blocks whose variation is small is reduced during clustering. In addition, K-modes and K-frequency algorithms belong to two-layer clustering strategy, the latter is more cohesive, because the K-frequency synthesizes the information of all attributes and is robust to singular points.

The larger the among-class distance is, the higher the separation is, the better the clustering result is. In Table I, K-frequency has the largest inter-class distance, followed by K-modes, not only shows that K-frequency clustering is better, but also shows that the two-tier cluster is more separable than single-tier.

TABLE II. COMPOSITE INDICATORS

Cluster methods	Dunn	Ch	Db	Ci
K-means	1.50e4	2.38e11	2.77	0.36
K-modes	1.77e4	2.27e11	3.09	0.15
K-frequency	1.61e4	2.60e11	2.08	0.64

The index, Dunn[8], represents the ratio of the maximum among-class distance to min-class distance. The larger the value is, the better the cluster result is. The index, Ch[9], represents the ratio of the sum of the distance between the

cluster center and all sample centers to the distance between all samples to each center. The bigger the value is, the better the cluster result is. The index, Db[10], represents the ratio of the distance from the sample point to the cluster center and the distance between the classes. The index, Ci[11], represents the ratio between the minimum and maximum values of the within-class distance differences.

In Table II, the indices Dunn, Ci indicate that K-modes are optimal. The indexes Ch and Db indicate that K-frequency is optimal. From Dunn, two-layer clustering is better than single-layer clustering.

ACKNOWLEDGMENT

This research was partially sponsored by the National Nature Science Foundation of China (Nos.61370113, 91546111,61672070 and 61272320), Beijing Municipal Natural Science Foundation (Nos.4152005 and 4162058), Key Projects of Beijing Municipal Education Commission (No KZ201610005009), Connotation Development Quota-international Cooperative Research Seed Grant (No. 007000514116520).

REFERENCES

- [1] R.Lawrence,"A Tutorial on Hidden Markov Models and Selected Application in Speech Recognition," in Proc.ICASSP,1989,77(2), PP:257-286.
- [2] R.Lawrence, B.Juang,"Fundamentals of Speech Recognition",Prentice Hall PTR,1993.
- [3] J.MacQueen,"Some methods for classification and analysis of multivariate observations" Proc of the 5th Symposium on Mathematical Statistics and Probability, Berkeley, 1967: 281-297.
- [4] Z.Huang,"Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values,"Data Mining & Knowledge Discovery, 1998, 2(3):283-304.
- [5] T.zhou, H.Lu,"Clustering algorithm research advances on data mining," Computer Engineering and Applications, 2012, 48(12): 100-111.
- [6] P.Lucey,J.F.Cohn,T. Kanade, J.Saragih,Z. Ambadar,I.Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression," Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010), San Francisco, USA, 94-101.
- [7] K.M. Lee,K.M.Lee,C.H Lee,"Statistical cluster validity indexes to consider cohesion and separation"2012 International conference on Fuzzy Theory and Its Applications (iFUZZY2012),2012,Pages: 228-232.
- [8] J.C.Dunn,"A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters,"Journal of Cybernetics,1973,3:32-57
- [9] R.B.Calinski,J.Harabasz,"A dendrite method for cluster analysis,"Communications in Statistics,1974,3(1):1-27
- [10] D.L.Davies,D.W.Bouldin,"A cluster separation measure,"IEEE Transaction on Pattern Analysis and Machine Intelligence,1979,2(PAMI-1):224-227
- [11] L.Hubert,J.Schultz,Quadratic assignment as a general data-analysis strategy,"British Journal of Mathematical and Statistical Psychology,1976,29(2):190-241